DOCUMENT RESUME

ED 279 691 TM 870 078

AUTHOR Peterson, Paul E.
TITLE Purposes of the National Assessment for Educational Progress.
PUB DATE 86
NOTE 30p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094. Bibliography (two pages) contains light, blurred type.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; *Educational Assessment; Educational Environment; Educational Testing; Elementary Secondary Education; *Evaluation Needs; Information Needs; Institutional Evaluation; *Measurement Objectives; *National Surveys; *School Effectiveness; Testing Problems; *Testing Programs; Test Use
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

The major purpose of the National Assessment of Educational Progress (NAEP) is to provide information useful for improving the quality of American elementary and secondary schools. NAEP has focused more on the measurement of academic achievement than on the need for information on learning contexts. Additional information is needed about why students perform as they do, and additional assessment of schools is desirable. NAEP has responded to its critics by treating and reporting all items separately and equally, countering the fear of creating a national curriculum. Recent efforts have strengthened test reliability and test validity, and earlier data has been reanalyzed to permit more reliable comparisons over time. NAEP data, although not as familiar as Scholastic Aptitude Test scores, have often been used to evaluate American education and achievement trends. However, more information is needed about learning contexts and school effectiveness: what works; causes of declines; and effects of peers, family, teachers, textbooks, curricula, finances, classroom organization, and special programs. Panel data comparing an individual's performance over the years while exploring contributing factors would be a costly but valuable data collection effort. NAEP might also be used to evaluate special education or compensatory programs, and national labor needs. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279691

Purposes of the National Assessment for Educational Progress

Paul E. Peterson

Brookings Institute

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. E. Peterson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

TM 870 078

Purposes of the National Assessment for

Educational Progress

by

Paul E. Peterson

The major purpose of the National Assessment for Educational Progress (NAEP) is to provide information useful for improving the quality of American elementary and secondary schools. To serve this purpose NAEP needs to obtain high quality information on both the educational attainments of American students and the contexts in which they learn. Over the past two decades NAEP has gathered increasingly high quality information on educational attainments, but it has yet to provide useful information on learning contexts. NAEP has told us much about what students have learned in reading, mathematics, writing, science, social studies and other subject areas, but it has produced very little information as to why students perform as they do. As a result, very little assessment of American schools has taken place or, indeed, is possible. The limits to NAEP's contributions are understandable for historical, political and institutional reasons, but until NAEP does more than report levels of achievement, it will make only modest contributions to the understanding or improvement of American Education.

In this paper we shall discuss NAEP's focus on measuring educational achievement, the need for information on learning contexts, and the ways in which a revised NAEP might contribute to

a variety of national purposes, including the collection of data on the effectiveness of national educational programs, strengthening national defense, and enhancing the quality of the nation's work force.

Measuring Educational Achievement

NAEP has always been concerned about obtaining meaningful measures of educational performance. Under the auspices of the Education Commission of the States, it sought to guard itself against the charge that the measures of student performance that it was obtaining were not appropriate to the curriculum the students were receiving or were not useful indicators of educational skills students should acquire. It obtained extensive assistance on test design from teachers, curriculum experts and educational psychologists. NAEP used a matrix sampling design that allowed it to ask a particularly large number of items covering a broad curricular range. It followed conservative reporting procedures that allowed readers to ascertain directly what percentage of the student population knew the answers to particular questions. Apart from the sheer summing up of the percentage correct of the answers given, it avoided the construction of scales, indices or any other measure of overall educational performance in a subject area.

Its caution in this regard was politically well taken. The idea of NAEP was initially criticized by the American Association of School Administrators as potentially dangerous. (American

Association of School Administrators, 1966). They expressed concern that an assessment might create national educational standards and, with it, a national curriculum. They also were concerned that an assessment might encourage teaching to the test and thereby distort school curricula. NAEP responded to these criticisms by developing a comprehensive instrument that treated all items separately and equally and by reporting data in a way that was immediately accessable to a lay audience.

Although this strategy was probably necessary in order to establish the concept of a quantitative national assessment, it raised serious methodological issues that eventually called for more sophisticated solutions. The validity and reliability of test instruments were not ascertained empirically. Simple summary measures did not permit accurate comparisons across time and among subgroups of the population. Tests were scheduled at various times during the school year, complicating comparisons over time. Age cohorts were defined inconsistently, and the timing of assessments was so haphazard that the progress of specific age cohorts could not be directly assessed.

In response to these criticisms the National Assessment was put up for competitive bid by the Department of Education in 1982. Management of NAEP was shifted from the Educational Commission of the States to the Educational Testing Service, which was well-known for its capacity to develop tests of aptitude and ability that were internally consistent and could partially predict subsequent performance. ETS used its considerable resources not

only to enhance the reliability and face validity of the tests it administered but also to reanalyze the data from earlier assessments in a way that permitted more reliable comparisons over time. It also has proposed a coherent schedule for assessments and developed consistent definitions of the populations to be sampled.

The success of the Educational Commission of the States in obtaining access to schools and the commitment to careful, precise measurement that is the hallmark of the Educational Testing Service are both laudable achievements. But it is worth considering whether limited resources are still being skewed too heavily in the direction of obtaining comprehensive measures of educational attainment at the cost of securing information about schools.

Accurate educational tests are especially important when attempting to assess the performance of individuals. In the case of the Scholastic Aptitude Test (SAT), the Graduate Record Examination, and other tests used for estimating future performances, it is crucially important that tests of achievement be highly reliable and valid at the individual level. The time, effort, and resources expended on developing an extensive battery of questions that have both face validity and internal consistency is critically important. Also, for these purposes, it is less important to gather information about learning contexts.

In the case of NAEP, however, the focus is not on the individual taking the test but the educational context within

which that individual is learning. We are interested in the quality of schools, not the abilities of specific individuals. Random errors at the individual level can be tolerated because at the group level these errors cancel one another, and data will still be accurate within an acceptably low margin of error. In recognition of this fact NAEP has used a matrix or spiral sample design that allows for a broad battery of test items to be administered within a one-hour examination period. Since no one student is asked the full battery of test items, it is impossible to ascertain accurately how well any one individual performed on the assessment. Yet standard statistical techniques are used to estimate with a low margin of error the performance of groups of students.

But even though NAEP recognizes that it is uninterested in the performances of specific individuals, it still has concentrated its resources on obtaining exact and extensive information on educational achievement in a variety of subjects. Instead of using score resources to increase our understanding of effective learning environments, it has tried to achieve a comprehensive understanding of the knowledge students have obtained.

NAEP data has nonetheless been used as an important source of information about the state of American education. In the early years of the program, to be sure, NAEP was given only passing attention by both the popular press and professional analysts. Absolute levels of student performances on various test items were

difficult to interpret. The public instead relied on other indicators of student performance, such as the Scholastic Aptitude Test, for their impressions of educational quality. But if NAEP was overlooked in the seventies, this changed sharply with the renewed interest in educational excellence that marked the eighties. The National Commission on Educational Excellence cast NAEP into the national limelight when it drew upon NAEP data--along with other information--to conclude that educational standards had deteriorated badly and that a renewal of commitment to educational excellence was urgently needed. Other nationwide commissions, citing similar findings, echoed the claims and recommendations of the Commission on Educational Excellence. Ever since, reform, quality, and improved teacher effectiveness have been key words shaping educational debates at national and state levels. Every national survey, including NAEP, has been scrutinized for signs of progress or decay.

NAEP made its own independent contribution to these discussions. NAEP is still not as popular an indicator of the state of American education as the SAT. It probably never will be, given the fact that the SAT is not a sample but is given to all students who register to take it and, as a result, most college educated Americans have taken the SAT and therefore have an intuitive sense of its meaning. NAEP data are nonetheless now taken quite seriously by those that give educational issues considered attention. Its data showing declining test scores among seventeen-year olds validated SAT data, which otherwise were

open to competing interpretations. NAEP also showed that the scores of nine-year olds remained quite constant or even improved in the seventies, helping to shift the attention of educational reformers to the high school. When NAEP found that it was the scores of the middle-class, white child that were declining the most, it strengthened the arguments of those who said more attention needs to be given to the education of average and gifted students. Even those who questioned the claim by national commissions that severe educational decline occurred in the 1970s turned to NAEP evidence (Congressional Budget Office, 1986; Peterson, 1983). The proper way to interpret educational achievement scores, these observers claimed, is not to view them simply as the outcome of the educational system at one point in time but instead to treat them as evidence about the accumulated educational experience of students over the course of their lives. When the data are viewed from this perspective, one finds that the cohorts of students attending school during the seventies and graduating in the eighties did better than the cohorts in school in the preceding decade.

Learning Contexts.

This debate over the meaning of NAEP data illustrates the difficulty of interpreting levels or changes in levels of educational achievement. If progress is occurring, what is it about the educational system that is working? If decline is evident, what are the causes of that decline? Are they to be found within families? in the peer group culture? or within

schools? If the causes are school-related, what are the key factors? Teachers effectiveness? Textbooks? Curricular co-ordination? Fiscal resources? Classroom organization? What effects do specialized programs, such as compensatory education, bilingual education, special education and vocational education, have on overall school effectiveness? What are the consequences of school reform efforts?

Information that will help answer these questions is information that will help us assess the progress of our educational system. It is the kind of information that will not only fuel educational debates but help pinpoint the problem and suggest appropriate courses of action. It is also the kind of information that could help fulfill a wide variety of national purposes.

Unfortunately, NAEP, as originally designed, was structured in ways that made answers to these kinds of questions difficult to obtain. School officials objected to comparisons among schools, districts and states, and, as a result, very little information was gathered on the contexts in which students were learning. Instead, information was reported by region, type of community, race, sex, and level of educational achievement. Collecting and reporting the data in this way eased access to the nation's schools on the part of the Education Commission of the States, but it did so at the cost of being able to tell us much that was significant about the progress of the nation's schools.

With the reconstitution of NAEP in 1982, the opportunity for improving information on learning contexts arose. The Educational Testing Service obtained in its 1983-84 assessment information on school characteristics from school principals and school teachers (Messick, Winter, 1985; Messick, November 1985). The initial report of this information is to be made in the fall of 1986, and at that time we will begin to have some information from NAEP on the way in which schools are affecting student performances. ETS reports that it is also making data tapes available to the educational research community so that other researchers and organizations will be able to do their own studies of school effectiveness (Messick, Beaton, and Lord, 1983).

There may well be important limitations to the studies and analyses that can be undertaken, however. The sampling design currently in use apparently does not generate a large enough data base to permit reasonably accurate estimates of the effects of schools variables on student achievement independent of the effects of background characteristics. Within the one-hour time limits of the test NAEP has concentrated on obtaining comprehensive, highly reliable data on student performance at the cost of obtaining usable information on learning contexts. (For a description of the sampling process, see Ibid). Unless major changes are made in sample and questionnaire design, this is likely to continue.

NAEP would also tell us much more about school effectiveness, if it collected information from a panel of

students as they proceeded through the educational system. Although panel data is costly to collect, it is exceedingly valuable. With panel data one can compare the performances of individuals at various stages of their educational career and explore the family, school, and other factors that contribute to changing levels of achievement from one testing period to the next. If the panel were continued into early adult lives (say, ages 21 - 25, the years recently surveyed by ETS), then analysts would also be able to see how educational experiences and levels of achievement affected future educational plans, occupational selection, employability, family life, and consumer choice.

Unfortunately, NAEP has thus far confined itself to the collection of cross-sectional information. To be sure, the current practice under ETS management is superior to the approach initially taken. According to the ETS plan, assessments will be conducted every four years so that we will in the future at least have information on the same cohort of students at ages 9, 13, and 17. Also, definition of age will be standardized, and the time of the year that the assessment takes place will be roughly the same. All of these steps are marked strides forward from the haphazard, inconsistent procedures that were instituted in the early years. Yet even with these improvements NAEP provides us with much less information about learning contexts than a panel survey can.

Panel surveys have their own problems, or course. The quality of the data tends to deteriorate over time. No matter how carefully individuals are tracked, attrition occurs and the sample

is no longer a random selection from the population that is being studied. These difficulties can be minimized if a portion of the sample (usually one-half) is renewed with each new wave of data collection. In this way one has both an acceptable cross-sectional estimate of population characteristics with each new assessment and a partial panel that permits an examination of the causes of educational change.

One way of organizing the examination of learning contexts is by state. SAT and ACT results are available on a state by state basis, and currently there are efforts underway to see whether more complete measures of educational achievement, such as provided by NAEP can be made available on a state by state basis. A NAEP sample for each of the fifty states could not be collected without greatly increasing the cost of the survey, however. There is some reason to doubt whether fifty state samples would be worth this cost. Student achievement varies more within states and districts than among them. The key school factors affecting student achievement are highly proximate to the student. They include teacher characteristics, classroom organization, curriculum, school morale, and principal leadership. While all of these factors may be indirectly affected by statewide policies, the casual effects are probably so complex and indirect that showing their impact on student achievement would prove virtually impossible. Certainly, a crude comparison of student performances by state will tell us very little about what specific state policies account for student achievement. Any state that wishes

to compare itself to national norms may conduct its own assessment; the fact that only one state is currently willing to pay the cost of such a survey suggests that states see little value in such information. The case for federal absorption of the cost is even less.

Instead of comparing states, the sample should be designed in such a way as to learn more about the factors affecting student achievement within schools. The collection of high-quality data on these learning contexts could make NAEP a more visible undertaking. In this regard, it is worth noting that two other surveys of American schools that did concentrate on school effectiveness have had far greater impact on policy discussions than NAEP has had. The first of these studies, known as the first Coleman report, was conducted under the Civil Rights Act of 1964 (Coleman et al., 1966). It compared the school resources available to black and white Americans. It found that educational resources that many had believed were important for school effectiveness had little effect on student achievement. Instead, the key factors were a child's family background and the influence of the child's peer group in school. The report justified the major effort at racial desegregation taking place in the late sixties and early 1970s, because it found that desegregation, by changing the racial composition of the school, could do more for the low-income, minority child's educational experience than any other change in school policy.

The second study, often referred to as Coleman II, was an analysis of the High School and Beyond survey commissioned by the National Center for Educational Statistics (Coleman, Hoffer and Kilgore, 1982). This study included for the first time extensive information on both public and private schools. Analysis of the data revealed that even when family background characteristics were controlled, the nonpublic schools were more effective educational institutions.

Both studies were highly controversial. Many scholars found limitations to the data and problems with the conclusions reached by the analysts. But unlike NAEP, these surveys illuminated important aspects of the educational process, and both have permanently changed the context for future research and policy discussions. Were NAEP to gather comparable information systematically from a panel of respondents on a regular basis, its likely impact would be much greater than either of these studies.

Other Potential Uses of NAEP

The main purpose of the NAEP is to provide information useful for improving the quality of American elementary and secondary schools. But NAEP may potentially be useful for a number of more particular national purposes. Suitably modified, NAEP data could be used to help ascertain the effectiveness of a number of national educational programs, such as compensatory education, vocational education, and special education. It may also assist the defense department in meeting its manpower needs as well as help the labor department estimate the need for and the effectiveness of employment and training programs.

Compensatory Education

There are three major federally funded educational programs--compensatory, vocational and special--whose effects can potentially be assessed using NAEP data. The educational outcomes of the largest of these programs, compensatory education, have not been studied in depth since the Sustaining Effects Study of 1976 (Kenayer, 1981). At the time that study was conducted NAEP data could not be used to assess program effectiveness. Although data was available by race and income, NAEP did not ascertain whether or not students in the sample had received compensatory services or not. Instead, the Sustaining Effects Study selected its own national sample of students receiving compensatory services as well as a comparison sample of students not receiving the services. This ten-year old study remains the country's most reliable and comprehensive data source on the program's effectiveness.

At the present time a major study of compensatory education has been undertaken at the request of Congress by the Office of Educational Research and Information, but due to a shortage of time and funding it was decided not to undertake another study of the educational effects of the program. The study also chose not to synchronize its data collections with NAEP because NAEP data would not be available in time for the Congressional deadline. Instead, the study has commissioned research looking at the administrative effectiveness of the program and the numbers and types of students receiving services.

In 1987 Congress will decide whether to authorize once again the compensatory education program. It is currently deliberating among a variety of possible alternatives. Some have suggested that Congress should require that funds be given only to schools in the most impoverished neighborhoods and then allow principals within these schools to develop programs for all the children in attendance. These commentators point out that the 1976 Sustaining Effects Study suggests that compensatory programs that pull eligible children out of the classroom for specific instruction by reading specialists or other specialized personnel are less effective than programs that increase the resources available to all the children within a compensatory school. The Department of Education has proposed an alternative reform; targeting the money on low-income pupils by issuing educational vouchers to their families. Still others have argued that current practices are successful and the deficiencies discovered in 1976 have since been corrected. However, none of these policy proposals can be based on recently collected empirical data on the impact of compensatory programs.

Undertaking a distinctive national study of compensatory education at each five-year interval the program comes before Congress for re-authorization has proven to be a difficult task. By the time administrative issues can be resolved, time and resources are often too limited to commission a national survey on the order of magnitude of the Sustaining Effects Study. Yet a study of that magnitude is often considered out of date by the

time the survey has been designed, the questionnaires distributed, the data collected and cleaned, and the analyses reported. But if NAEP were to collect high quality data on learning contexts, it could, by adding a few additonal relevant items, be able to provide appropriate information on the effectiveness of this large, federally funded program.

One important limit on the data potentially collected by NAEP needs to be acknowledged. NAEP data on students participating in compensatory education programs could not be compared to a randomly selected control group that did not participate. Some innovative programs, such as the negative income tax, manpower training and state workfare programs, have been evaluated using experimental design data. Since these experimental design studies randomly assign eligible recipients of public services to participant and control groups, they can identify the effects of the policy experiment, controlling for all other variables. The only way in which this can be approximated with NAEP data is by using regression techniques that statistically control for the variety of family background and school variable that are thought to be relevant. Although such a research design can be quite sophisticated, it can never determine conclusively whether or not a particular educational practice being studied <u>really</u> accounted for the differences being observed.

The Coleman II study mentioned earlier became entangled in exactly this kind of methodological controversy (Coleman, Hoffer and Kilgore, 1982). The authors reported that students attending

nonpublic schools performed better on a variety of tests of educational achievement, even after family background characteristics were taken into account. Critics replied that the study did not take into account all of the relevant family background characteristics. Controlling for family income and educational levels was insufficient. The family commitment to education expressed by the fact that they were willing to pay for a private education was not taken into account.

Up until now NAEP has escaped these kinds of controversies, and perhaps, as a result, it has established a secure political base. But the cost has been high; NAEP data thus far tell us little about learning contexts and without that data NAEP can make only modest contributions.

Vocational Education

Congress passed a new Vocational Education Act in 1985 and in that legislation it called for a National Assessment of Vocational Education. The planning for that assessment by the Department of Education is now underway, and an overall design conference is scheduled for early September, 1986. One of the issues expected to emerge at the design conference is whether or not data should be collected on the relationship between curricular choice and educational attainment. The study also seeks to ascertain the effect of the educational reform movement on vocational education's place in the high school curriculum. Still a third topic for which data is being collected is on the vocational training of handicapped students. On all three topics

consideration is being given to the possibility of drawing either directly upon NAEP data or upon the sampling frame that NAEP is using. Some of the advantages of co-ordinating the vocational education study with NAEP is that the sampling frame is already in place, participating schools have already agreed to co-operate, and approval for data collection is more easily obtained from the Federal Educational Data Acquisitions Council.

These are the first efforts to use NAEP resources to evaluate federal vocational education efforts. Information on the effects of vocactional education have in the past depended heavily upon the National Longitudinal Study of 1972 and the High School and Beyond Study, which began in 1980. The latter study is collecting information during the sophomore year, the senior year and every two years after the completion of high school. Thus, the data set allows for an examination of the transition of school to work and for an examination of the effects of education on later life experience. If NAEP is to be an equally valuable tool for studying the effects of varying kinds of curricular experiences, such as vocational education, it is necessary that panel data be collected. This method would be especially appropriate for any future 21 - 25 age cohort to be included in NAEP. If the young adults interviewed in their early twenties are the same individuals as those interviewed in the closing years of high school, one could then analyze effects of curriculum, educational achievement, and other school characteristics on subsequent life experiences.

We know very little about the practical value of scoring high on various tests of educational achievement, though most people believe improved educational achievement enhances national defense, worker productivity, domestic prosperity, consumer awareness, family stability and political participation. The evidence that education has these effects comes mainly from studies that show that individuals that remain in school over longer periods of time are subsequently more successful in their occupational and personal lives.

These findings have been questioned, however, for the selectivity bias inherent in the data being used. Those who remain in school longer may have personal attributes that account for their subsequent achievements quite independent of their educational experiences. At most, schools and colleges give them the credentials they need for later success. This argument is supported by research that shows that verbal ability, independent of family background and years of education, accounts for little of the variance in the personal incomes of adults.

But these findings, which question the value of educational achievement, are themselves based on very inadequate data sets. We still do not have high-quality panel data that allows the analyst to link school experiences to later life experiences directly and to ascertain the degree to which curriculum, educational attainment, and other school factors actually shape subsequent life experiences.

The value of the 21 - 25 year old NAEP sample is thus not so much in identifying the level of educational attainment of this age cohort. At this point one is no longer assessing even in an approximate sense the quality of educational instruction in the nation's elementary and secondary schools. The value of this data collection is that it allows the analyst to see the connections between educational experiences and the assumption of adult roles in society. The educational testing service is planning to collect any future data from the 21 - 25 year old cohort that they tested in high school. That is a small step in the right direction because the sample will at least be drawn from the same population sampled in school. But the utility of the data would be greatly enhanced if the survey included the same individuals for whom data is available on educational experiences and achievements.

Special Education

Education for the handicapped is the second largest of all federal education programs. The principal statutory basis for the program is the Education of All Handicapped Children Act of 1974, commonly known as P. L. 94-142. This law gives every child the right to an appropriate education between the years 3 and 21; it requires that school districts develop individualized educational plans for each child, hold due process hearings in cases where parents and school officials cannot agree on a plan; institute child-find procedures which insure that children in need of specialized services are identified; and arrange for all related

services the child may need. The costs to local authorities of instituting this important federal initiative have greatly exceeded the amount the federal government is currently allocating for the programs. Groups representing handicapped children and their families report that the quality and quantity of educational services for the handicapped have greatly improved as a result of this legislation. However, no national survey of the impact of this program has ever been conducted.

Currently, the Department of Education is carrying out a survey of the non-educational services available to handicapped students ages 13, 15, 17, 19, and 21. This survey of the parents of some 8,000 students will be conducted in two stages, four years apart, so that data will become available on the processes of adjustment that are required as handicapped students move out of the secondary educational system into a setting appropriate to their adult status. This study will also gather some information on the educational experiences of the child.

It is not clear why this data is not being collected within the sampling frame provided by NAEP. It is probably due to the fact that historically NAEP provided very little information on handicapped students. Although the Educational Testing Service is now apparently gathering information on persons excluded from educational testing, NAEP'S period of transition has overlapped the development of the design for the survey of the handicapped.

In the future, it should be possible to build on the improved data gathering capacities of NAEP so that its data could

be made available for assessing progress in special education. By oversampling this population, estimates could be obtained on the numbers of students receiving special education services, the nature of the services being received, the extent of mainstreaming that is occurring, and, for as many as 80 per cent of the students, the level of educational attainment being realized. Eventually, it might be possible to make special provision for the testing of handicapped children excluded from NAEP. The value of this information would increase, if panel data could be gathered on handicapped student, allowing for the identification of transitions out of special education programs and into adult roles.

National Defense

A disjunction between the increasing sophistication of military equipment and the potential decline in the quality of military personnel has created a new sense of urgency concerning the educational qualifications necessary to maintain the national defense. In the years immediately following the replacement of the draft by an all-volunteer army, the quality of military personnel deteriorated noticeably. But in the 1980s much of the lost ground has been regained, and today the ability of military personnel, as measured by receipt of the high school diploma and performance on the Armed Forces Qualifying Test (AFQT), is as high as it was when the draft was still in effect. The major concern is for the future, as the size of the age cohort available for military service is about to decline and its demographic composition about to change (Binkin, 1986).

As a sign of this concern, personnel from the armed services recently attended a briefing given by NAEP staff members. They discussed not only the recent gains by high school students, but also the worrisome signs of decline in the elementary years. Although interested in information on these trends, the armed forces are little different from any other potential employer. Their expectation of NAEP is mainly that information be used in such a way as to improve school effectiveness and, as a result, the capabilities of the next generation.

More particular interests of the armed forces are served by its own educational tests. In addition to the AFQT, the individual services "have developed composite sets of subtests, designed to predict training success in particular job clusters, such as electronics repair and mechanical maintenance. . . . Although scores on composite tests are used primarily to assess vocational aptitude, the armed forces also use them to supplement AFQT scores and educational attainment in determining basic eligibility for entry into certain services. To qualify for enlistment in the Army in 1985, for example, a high school graduate was required not only to place above the 15th percentile on the AFQT but also to score 85 or above on one of the ten Army occupational-area aptitude composites" (Binkin, 1986, pp. 17-18).

As this passage suggests, the armed forces use tests of aptitude and achievement in ways very similar to the uses made of them by institutions of further education and by other employers. They are interested in obtaining information on the performances

of individuals so as to determine their fitness for a military career. They are also valuable as recruitment devices, because they give the armed forces an opportunity to explain to recruits the educational, occupational and other benefits of a military training. NAEP could not serve as a substitute for this very specific information on potential recruits.

Suitably revised, NAEP data might be of use for evaluating the military's own extensive educational programs, however. At the present time we know very little about the effectiveness of military training for later civilian careers. Yet the armed forces has an extensive, carefully designed educational program that has served a disproportionately large percentage of minority Americans. Were panel data collected on high school students and again in their adult years, it would be possible to ascertain the effectiveness of various kinds of military experience on civilian employment and earnings.

Labor Market Research

The Department of Labor is interested in ascertaining and improving the quality of the nation's workforce. This interest is especially great at a time when the United States' competitive advantage depends on building an increasingly skilled workforce. However, the labor department's current interest in NAEP data is limited by the fact that NAEP does not allow for an examination of the connection between educational experience and later work experiences. If the data were suitably revised along the lines discussed in the section of this paper devoted to vocational

education, NAEP information could be used to study the effects of education on performance in the labor market. Also, the data could be used to help study the effectiveness of manpower training programs, workfare policies, and other federal interventions to assist marginal members of the economy. However, it would probably by necessary to oversample subgroups of the population at risk of unemployment and underemployment. The High School and Beyond Study as well as the planned National Longitudinal Study of the Class of 1988 are obtaining information of this type.

Conclusions

NAEP has succeeded in convincing the American public of the value of periodic assessments of the educational achievements of students at various points during their educational career. It has also developed recently an increasingly sophisticated sampling frame and test design that allows for reliable comparisons over time and across subgroups of the population. Some information on learning contexts is also being gathered, and we will soon learn more about the utility of this information. However, the focus of NAEP has thus far been concentrated on measuring educational outcomes with great precision. A price has been paid for this precision. Very little of an extremely limited research budget has been used to gather information on learning contexts. As a result we know quite a bit about overall student educational progress (or decline), but hardly anything about school effectiveness. NAEP will be more valuable in the future if it

designs its sample, tests and other data collection efforts in such a way as to gather more information about the schools in which students are learning and more information about the same students as they progress through the educational system and into adult roles. If this is done, NAEP will be much more useful for helping design improvements in our educational system, for evaluating the effectiveness of federal programs, for informing the Departments of Labor on the effectiveness of its programs, and aiding in the national defense.

SELECTED BIBLIOGRAPHY

American Association of School Administrators. National Educational Assessment: Pro and Con. Washington, D.C.: National Education Association, 1966

Anderson, Beverly L. "State Testing and the Educational Measurement Community: Friends or Foes?" Educational Measurement: Issues and Practice, vol. 4, (Summer, 1985)

Anrig, Gregory, and Archie Lapointe. "NAEP: The Nation's Report Card," Phi Delta Kappan, vol. 65, (September 1983)

Binkin, Martin. Military Technology and Defense Manpower. Brookings, 1986

Chubb, John E., and Terry M. Moe. "Politics, Markets, and the Organization of Schools," Discussion Paper 1 (Brookings, 1986)

Coleman, James S., and others. Equality of Educational Opportunity. U.S. Department of Health, Education, and Welfare, Office of Education (GPO, 1966)

_____, Thomas Hoffer, and Sally Kilgore. High School Achievement. New York: Basic Books, 1982

Comptroller General of the United States, Report to the Congress: The National Assessment of Educational Progress: Its Results Need to Be Made More Useful, 1976

Jencks, C.S., M. Smith, H. Ackland, M.J. Bane, D. Cohen, H. Gintis, B. Heyns, and S. Michelson. Inequality: A Reassessment of the Effect of Family and Schooling in America. Santa Monica, CA: The Rand Corporation, 1973

Kay, Brian. "National Assessment of Educational Progress in the United States," Trends in Education, (1976)

Kennedy, Mary M., Richard K. Jung, and Martin E. Orland. Poverty, Achievement and the Distribution of Compensatory Education Services. U.S. Department of Education, Office of Educational Research and Improvement (GPO, 1986)

Kenoyer, C.E., D.M. Cooper, D.E. Saxton, and R. Hoepfner. The Effects of Discontinuing Compensatory Education Services. Technical Report No. 11 from the Study of the Sustaining Effects of Compensatory Education on Basic Skills. Santa Monica, CA: Systems Development Corporation, 1981

Lapointe, Archie E., and Stephen L. Koffler. "Your Standards or Mine?" Educational Researcher, vol. 11, (December, 1982)

Layton, Donald H. "ECS at 20: New Vitality and New Possibilities," Phi Delta Kappan, vol. 67, (December, 1985)

Maxwell, John C. "National Assessment of Writing: Useless and Uninteresting?" English Journal, vol. 62, (December 1973)

Messick, Samuel. "Progress Toward Standards as Standards for Progress: A Potential Role for NAEP," Educational Measurement: Issues and Practice, (Winter 1985)

_____. "Response to Changing Assessment Needs: Redesign of the National Assessment of Educational Progress," American Journal of Education, vol.94, (November 1985)

_____, Albert Beaton, and Frederic Lord. A New Design For a New Era. Princeton, NJ: Educational Testing Service, 1983

National Institute of Education. Evaluating Compensatory Education: An Interim Report on the NIE Compensatory Education Study. Washington, D.C.: NIE, 1976

Peterson, Paul. Making the Grade. New York: Twentieth Century Fund, 1983

_____. "Did the Education Commissions Say Anything?" The Brookings Review, vol.2, (Winter, 1983)

Power, Colin, and Robert Wood. "National Assessment: A Review of Programs in Australia, the United Kingdom, and the United States," Comparative Education Review, vol. 28, (August, 1984)

Sebring, Penny A., and Robert F. Boruch. "How is National Assessment of Educational Progress Used?" Educational Measurement: Issues and Practice, vol. 2, (Spring 1983)

Shane, Harold G. "Measuring Educational Quality: Harold G. Shane Interviews Willard Wirtz on the National Assessment of Educational Progress," Phi Delta Kappan, (December 1981)

Tyler, Ralph. "Comment on Lapointe and Koffler, "Educational Researcher, vol. 11, (December, 1982)

_____. "Educational Assessment, Standards, and Quality: Can We Have One Without the Others?" Educational Measurement: Issues and Practice. vol. 2, (Summer, 1983)

Wiley, David E. "Improving Policy Development," New Directions for Testing and Measurement, vol. 10, (1981)

Wirtz, Willard and Archie Lapointe. "Measuring the Quality of Education: A Report on Assessing Educational Progress." Washington, D.C.: Wirtz and Lapointe, 1982.

Womer, Frank B., and Margorie M. Mastie, "How Will National Assessment Change American Education?" Phi Delta Kappan, vol. 53, (October 1971)